ELSEVIER

Journal of Cereal Science 47 (2008) 429–437

Journal of CEREAL SCIENCE

www.elsevier.com/locate/jcs

Isolation and characterization of a novel variant of HMW glutenin subunit gene from the St genome of *Pseudoroegneria stipifolia*

Z.X. Li[a,c], X.Q. Zhang[a,*], H.G. Zhang[a,b], S.H. Cao[a], D.W. Wang[a,**], S.T. Hao[a], L.H. Li[d], H.J. Li[d], X.P. Wang[a]

[a]*State Key Laboratory of Plant Cell and Chromosome Engineering, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China*

[b]*Northwest Institute of Plateau Biology, Chinese Academy of Sciences, Xining 810008, China*

[c]*Graduate University of Chinese Academy of Sciences, Beijing 100039, China*

[d]*Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing 100081, China*

Received 9 December 2006; received in revised form 16 May 2007; accepted 23 May 2007

Abstract

The x- and y-type high molecular weight (HMW) glutenin subunits are conserved seed storage proteins in wheat and related species. Here we describe investigations on the HMW glutenin subunits from several *Pseudoroegneria* accessions. The electrophoretic mobilities of the HMW glutenin subunits from *Pd. stipifolia*, *Pd. tauri* and *Pd. strigosa* were much faster than those of orthologous wheat subunits, indicating that their protein size may be smaller than that of wheat subunits. The coding sequence of the Glu-1St1 subunit (encoded by the *Pseudoroegneria stipifolia* accession PI325181) was isolated, and found to represent the native open reading frame (ORF) by *in vitro* expression. The deduced amino acid sequence of Glu-1St1 matched with that determined from the native subunit by mass spectrometric analysis. The domain organization in Glu-1St1 showed high similarity with that of typical HMW glutenin subunits. However, Glu-1St1 exhibited several distinct characteristics. First, the length of its repetitive domain was substantially smaller than that of conventional subunits, which explains its much faster electrophoretic mobility in SDS-PAGE. Second, although the N-terminal domain of Glu-1St1 resembled that of y-type subunit, its C-terminal domain was more similar to that of x-type subunit. Third, the N- and C-terminal domains of Glu-1St1 shared conserved features with those of barley D-hordein, but the repeat motifs and the organization of its repetitive domain were more similar to those of HMW glutenin subunits than to D-hordein. We conclude that Glu-1St1 is a novel variant of HMW glutenin subunits. The analysis of Glu-1St1 may provide new insight into the evolution of HMW glutenin subunits in *Triticeae* species.

Keywords: *Pseudoroegneria stipifolia*; HMW glutenin subunit; D-hordein; Wheat; Phylogenetic analysis; Evolution

1. Introduction

High molecular weight (HMW) glutenin subunits are important determinants of the baking quality of bread wheat (*Triticum aestivum* L.) (Payne, 1987; Shewry and Halford, 2002). There are two subfamilies of HMW glutenin subunits, known as x- and y-types arising from a gene duplication event that predates the divergence of the A, B, and D genomes of cultivated wheat (Gu et al., 2003). In hexaploid wheat, HMW glutenin subunits are encoded by genes contained in the *Glu-A1*, *Glu-B1* and *Glu-D1* loci located on the homoeologous group one chromosomes

Abbreviations: HMW-GS, high-molecular-weight glutenin subunit; IPTG, isopropyl b-D-thiogalactopyranoside; kDa, kilo Dalton; MALDI-TOF-MS, matrix assisted laser desorption ionization time-of-flight mass spectrometry; MYA, million years ago; ORF, open reading frame; PCR, polymerase chain reaction; SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis

*Corresponding author. Tel.: +86 10 64873482; fax: +86 10 64873482.

**Also for Correspondence.

E-mail addresses: xqzhang@genetics.ac.cn (X.Q. Zhang), dwwang@genetics.ac.cn (D.W. Wang).

0733-5210/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jcs.2007.05.010

(Lawrence and Shepherd, 1981; Payne, 1987; Thompson et al., 1983). Theoretically, there should be six expressed HMW glutenin subunits in hexaploid wheat. However, the compositions of HMW glutenin subunits often vary among bread wheat cultivars due to gene silencing and allelic variation. Usually, three to five subunits are expressed in individual varieties (Payne and Lawrence, 1983). The primary structure of a HMW glutenin subunit is composed of four regions, the signal peptide (removed from mature protein), N- and C-terminal domains, and a central repetitive region (Shewry et al., 1995). In the N- and C-terminal regions, there exist cysteine residues that are highly conserved in both numbers and positions. The repetitive domain predominantly consists of repeats encoding tri-, hexa- and nonapeptides (Shewry et al., 1995). Both the conserved cysteine residues and the size of the repetitive domain contribute to the higher order structure of HMW glutenin subunits. The former are involved in the formation of disulphide bonds, the latter may initiate intermolecular interactions through hydrogen bonds (Shewry et al., 1995). Variations in the numbers of repeated peptide motifs often lead to changes in the length of the repetitive domains, which is the main cause for size differences among different subunits (Shewry et al., 1995). The open reading frames (ORFs) for HMW glutenin subunit genes are generally 1.8–2.5 kb, and the calculated molecular mass of these subunits ranges from 65 to 90 kDa (Shewry et al., 1995).

Orthologous HMW glutenin subunits have been found in many *Triticeae* grasses including various *Aegilops* species and rye (De Bustos et al., 2001; De Bustos and Jouve, 2003; Liu et al., 2003; Wan et al., 2002; William et al., 1993). In barley (*Hordeum vulgare* L.), the seed storage protein D-hordein is structurally related to HMW glutenin subunits of wheat (Halford et al., 1992). The N-terminal domain of D-hordein is more similar to that of typical y-type HMW glutenin subunits, whereas its C-terminal domain exhibits higher resemblance to that of typical x-type HMW glutenin subunits (Gu et al., 2003). However, the central repetitive region of D-hordein shows multiple differences from that of typical HMW glutenin subunits in both domain organization and repeat motifs (Gu et al., 2003). The D-hordein central domain is composed of two repetitive regions separated by a non-repetitive and cysteine-containing fragment that is unique to D-hordein. The repeated motifs contained in the first repetitive region (PGQGQQ, PGQGQQGYYPSATSPQQ) bear high resemblance to those found in the repetitive domain of HMW glutenin subunits. However, the main repeated motif in the second repetitive region of D-hordein (PHQGQQTTVS) has no equivalent in the HMW glutenin subunits characterized thus far (Gu et al., 2003).

Since HMW glutenin subunit genes display similar temporal and spatial patterns of expression (Thomas and Flavell, 1990), they must have conserved 5′ flanking regulatory elements. Previous studies have indicated that HMW glutenin subunit genes contain a major regulatory element 5′-GTTTTGCAAAGCTCCAATTGCTCCTTG-CTTATCCAATATT-3′ in their promoter regions (Thomas and Flavell, 1990). The location of this element is highly conserved in all HMW glutenin gene promoters reported to date, beginning at positions −185 to −189 (Shewry et al., 1999). In addition, HMW glutenin subunit genes also have some common *cis*-acting elements in their promoter regions, such as E box (TGTAAA), N box (TGAGTCA), G box (TTACGTGG) and TATA box (TATAAAA) (Thomas and Flavell, 1990).

In contrast to the studies described above, there is still little information on the HMW glutenin subunits and their coding genes in the St genome, which is contained in a wide range of *Triticeae* species (McMillan and Sun, 2004; Xu and Ban, 2004). The diploid and tetraploid *Pseudoroegneria* species have been considered to be the donor of St genome in many natural intergeneric *Triticeae* hybrids (Jensen et al., 1990; Lu, 1994; McMillan and Sun, 2004; Xu and Ban, 2004). In this work, we have investigated the HMW glutenin subunits in several diploid and tetraploid *Pseudoroegneria* accessions, with the Glu-1St1 subunit (expressed in the diploid *Pd. stipifolia* accession PI325181) being analyzed in more detail. The results of our investigations are reported in the following sections, together with the discussions on the unusual structural features of Glu-1St1 and the new insight gained into the evolution of HMW glutenin subunits based on the analysis of Glu-1St1.

2. Materials and methods

2.1. Plant materials

Seeds of two diploid *Pd. stipifolia* accessions (PI531750 and PI325181, $2n = 14$, StSt), one tetraploid *Pd. tauri* accession (PI380650, $2n = 28$, StStPP) and one tetraploid *Pd. strigosa* accession (PI531752, $2n = 28$, StStStSt) were obtained from the Institute of Crop Sciences of the Chinese Academy of Agricultural Sciences, Beijing, China. These accessions were originally obtained from the United States Department of Agriculture (USDA). The HMW glutenin subunits from the bread wheat variety Chinese Spring (1Bx7 + 1By8, 1Dx2 + 1Dy12) were used as electrophoretic mobility standards in SDS-PAGE experiments.

2.2. SDS-page and protein blot analysis

HMW glutenin subunits were extracted from the seeds and were separated using SDS-PAGE as previously described (Deng and Zhang, 2004). At least five individual seeds were examined to ascertain the composition of HMW glutenin subunits in a given accession. A polyclonal antibody recognizing both x- and y-type HMW glutenin subunits of wheat was used in protein blot analysis according to the method published previously (Wan et al., 2002).

2.3. PCR amplification

Genomic DNA was extracted from the young seedlings of *Pd. stipifolia* accession PI325181 as described previously (Wan et al., 2002). To amplify the ORF of HMW glutenin gene from PI325181 by genomic PCR, two degenerate primers P1 (5′-ATCACCCACAACACCGAGCA-3′G) and P2 (5′-CTATCACTGGCTA/GGCCGACAATGCG-3′), capable of amplifying the complete coding sequences of HMW glutenin genes from wheat related species (Liu et al., 2003), were used. To amplify the 5′-flanking sequence of HMW glutenin subunit gene by genomic PCR, the degenerate primers Pp1 (5′-AGGGAAAGACAATGGA-CATG-3′) and Pp2 (5′-TAGTTG/TCCC/TAGAGGCCT-CACCTTC-3′) were designed and used. PCR amplification was carried out in 20 μL reaction mixture containing 100 ng of DNA, 10 μL of 2 × GC (to amplify DNA fragment with high GC content) buffer, 400 μM each dNTPs, 0.1 μM each primers, and 1 U of LA (Long and Accurate) Taq DNA polymerase (TaKaRa, Dalian, China). Reaction mixture was denatured for 4 min at 95 °C prior to 33 cycles of 1 min at 95 °C, 50 s at 62 °C, and 3.5 min at 72 °C. The final extension step was 10 min at 72 °C. Amplified products were separated in 1% agarose gel, and the anticipated band was excised from the agarose gel and purified using a DNA gel extraction kit (Sangon, Shanghai, China). DNA fragment was ligated into the pGEM-T vector (Promega, USA) and then sequenced in an Applied Biosystem 3730 DNA Analyzer by a commercial company (Bioasia, Beijing, China).

2.4. Bacterial expression of the cloned ORF of Glu-1St1

To confirm that the cloned DNA fragment was an accurate representation of the coding sequence of Glu-1St1 subunit, bacterial expression experiments were conducted. The cloned ORF fragment was re-amplified using the primers P3 (5′-ATTCATATGGAAGGTGAGGCCTCT-GG-3′) and P4 (5′-CTAGAATTCTATCACTGGCTG-GCCGA-3′) to remove the signal peptide and to add restriction enzyme sites (*Nde*I and *Eco*RI) for subsequent cloning work. The resulting fragment was cloned into the bacterial expression vector pET-30a (Invitrogen, USA). The induction of the pET construct by IPTG (isopropyl-β-D-thiogalactopyranoside, Sigma) was carried out as described previously (Wan et al., 2002). The electrophoretic mobility of the overexpressed protein in the induced bacterial cells was compared to that of the native Glu-1St1 subunit extracted from the seeds using SDS-PAGE.

2.5. MALDI-TOF-MS analysis of native Glu-1St1

The seed protein sample extracted from PI325181 was separated using 10% SDS-PAGE with Tris–Glycine-SDS running buffer. The native Glu-1St1 band was excised from the gel, followed by in gel digestion with trypsin (Kumarathasan et al., 2005). A sample of the digested protein (0.5 μL) was subject to MS analysis in an Autoflex MALDI-TOF mass spectrometer (Bruker Daltonics, MA, USA) as described previously (Cunsolo et al., 2002, 2003; Sun et al., 2006). Three independent MS experiments were performed to test the reproducibility of the results.

2.6. Sequence analysis and evolutionary investigations

To investigate the phylogenetic relationships of the *Glu-1St1* glutenin subunit gene with previously characterized *Glu-1-1* alleles (represented by *1Ax*, *1Bx7*, *1Dx2*, *1Cx* and *1Ux*), *Glu-1-2* alleles (represented by *1Ay*, *1By9*, *1Dy12*, 1Cy and 1Uy), *Glu-1R* alleles from rye and *D-hordein* from barley, multiple alignments were created with the amino acid or nucleotide sequences using the Clustal W program (Thompson et al., 1994). The alignment files were converted to MEGA format for building the phylogenetic trees as detailed previously (Li et al., 2004; Nei and Kumar, 2000). The divergence times of *Glu-1St1* from orthologous HMW glutenin subunit alleles in wheat and rye and from barley *D-hordein* were estimated using the nucleotide sequences of the N- plus C-terminal domains of the compared genes. The methods involved were detailed previously (Li et al., 2004; Sanderson, 1998), with the average nucleotide substitution rate of 6.5×10^{-9} per site per year, originally calculated for barley *ADH* genes (Gaut et al., 1996), used during the estimations. For generating multiple sequence alignments, appropriate DNA or protein sequences of *1Ax*, *1Bx7*, *1Dx2*, *1Cx*, *1Ux*, *1Ay*, *1By9*, *1Dy12*, *1Cy*, *1Uy*, *1Rx*, *1Ry*, and *D-hordein* genes were retrieved from the GenBank (http://www.ncbi.nlm.nih.gov/). The accession numbers for these genes were AF145590, X13927, X03346, AF476959, AF476961, AJ306977, X61026, X03041, AF476960, AF476962, AJ314782, AJ314780 and AY268139, respectively.

3. Results

3.1. HMW glutenin subunits in Pseudoroegneria accessions

SDS-PAGE analysis using seed protein extracts revealed the presence of one putative HMW glutenin subunit in each of the two *Pd. stipifolia* accessions (PI531750 and PI325181, Fig. 1A, lanes 2 and 3). By contrast, the HMW glutenin subunit compositions in the tetraploid *Pseudoroegneria* accessions were more complex. In the *Pd. tauri* accession PI380650, two subunits with very close electrophoretic mobilities were observed (Fig. 1A, lane 4), whereas four subunits were found in the *Pd. strigosa* accession PI531752 (Fig. 1A, lane 5). The HMW glutenin subunits from the *Pseudoroegneria* accessions examined in this work all shared one common feature that their electrophoretic mobilities were much faster than those exhibited by their counterparts from Chinese Spring (Fig. 1A, lane 1). Protein blot experiments demonstrated that the subunits from *Pseudoroegneria* accessions all reacted strongly with the polyclonal antibody specific for

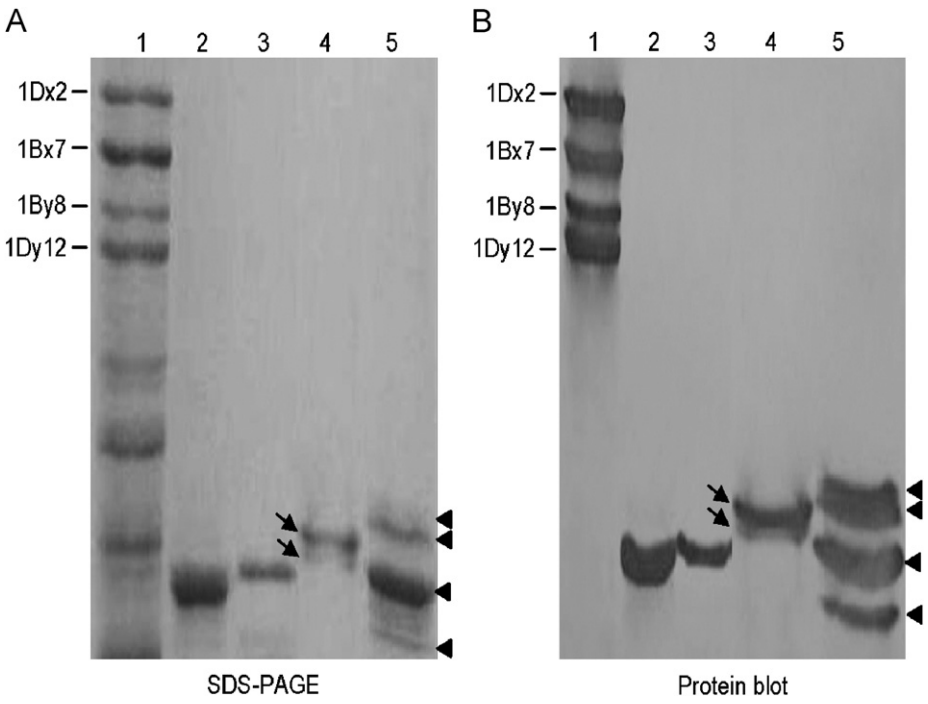

Fig. 1. SDS-PAGE (A) and protein blot (B) analyses of the HMW glutenin subunits expressed in *Pseudoroegneria* species. In both (A) and (B), the HMW glutenin subunits from the hexaploid wheat variety Chinese Spring (1Dx2, 1Bx7, 1By8, and 1Dy12, lane 1 were used as the controls. The HMW glutenin subunits expressed in the *Pd. stipifolia* accessions PI531750 and PI325181 were contained in lanes 2 and 3, respectively. Two subunits with very close electrophoretic mobilities (indicated by arrows in lane 4 were found for the *Pd. tauri* accession PI380650, whereas four subunits (indicated by arrowheads in lane 5 were detected for the *Pd. strigosa* accession PI531752. The results shown in (A) and (B) are representative of five separate sets of experiments.

HMW glutenin subunits (Fig. 1B), thereby confirming their close structural relationships with HMW glutenin subunits. The results depicted in Fig. 1B also confirmed the number of expressed subunits revealed by SDS-PAGE for each of the four *Pseudoroegneria* accessions (Fig. 1A).

3.2. Analysis of Glu-1St1 ORF and deduced amino acid sequence

The unusual electrophoretic mobilities displayed by *Pseudoroegneria* subunits prompted us to investigate their genomic coding sequences. To simplify our molecular analysis, we chose the diploid *Pd. stipifolia* accession PI325181 as the representative because only one subunit (tentatively named as Glu-1St1) was found to be expressed in this accession (Fig. 1). In genomic PCR reactions using the primers P1 and P2, a single PCR fragment was specifically amplified from the genomic DNA sample of PI325181 (Fig. 2A, indicated by arrowhead). The PCR product was cloned, and confirmed to be highly similar to HMW glutenin subunit gene coding sequences reported previously by DNA sequencing and nucleotide sequence comparisons. An intact ORF containing 1227 nucleotides was identified in the cloned fragment. When expressed in the bacterial cells, the electrophoretic mobility of the overexpressed polypeptide (Fig. 2B, marked by asterisk) was identical to that of the native Glu-1St1 subunit extracted from PI325181 seeds (Fig. 2B, arrowed), indicating that the cloned ORF is an accurate representation of the native *Glu-1St1* coding sequence.

The amino acid sequence deduced from the cloned *Glu-1St1* ORF contained 407 amino acid residues (Fig. 3). To verify the deduced sequence, three independent MALDI-TOF-MS experiments were conducted using the native Glu-1St1 subunit. The molecular mass of 12 peptides could be reliably determined (Table 1). The mass values of the 12 peptides matched with those predicted based on the deduced amino acid sequence of Glu-1St1. These peptides encompassed almost the complete deduced amino acid sequence of Glu-1St1 (Fig. 3, underlined), suggesting that the deduced amino acid sequence agrees with that of native Glu-1St1.

Amino acid sequence comparisons indicated that, like typical HMW glutenin subunits, the Glu-1St1 subunit protein contained a signal peptide, a N-terminal domain, a central repetitive domain, and a C-terminal domain (Fig. 3). The signal peptide of Glu-1St1 contained 21 amino acid residues (Fig. 3), being the same as that recorded for published HMW glutenin subunits (Li et al., 2004). In the N-terminal domain, Glu-1St1 had 105 amino acids, among which five were cysteine residues (Fig. 4A). Both the size and the number and positions of the cysteine residues in Glu-1St1 N-terminal domain resembled highly those of typical y-type subunits and D-hordein (Fig. 4A). Moreover, the last six residues of Glu-1St1 N-terminal domain (SSQTVQ, Fig. 4A, boxed region) were identical to those found in typical y-type subunits. However, the

C-terminal domain of Glu-1St1 was more similar to that of typical x-type subunits and D-hordein than to the corresponding domain in typical y-type subunits (Fig. 4B). The amino acid sequence element LAAQL-PAMCRL (spanning residues 387–397 in Glu-1St1 C-terminal domain, Fig. 3) was completely conserved among Glu-1St1, typical x-type subunits and D-hordein (Fig. 4B, boxed region).

The central repetitive domain of Glu-1St1 had only 239 amino acid residues, which was much shorter than that recorded for typical HMW glutenin subunits (Fig. 4C). This led to the calculated molecular mass of Glu-1St1 (44 kDa) being much lower than that found for typical HMW glutenin subunits (65–90 kDa, Shewry et al., 1995). Glu-1St1 repetitive domain contained 25 hexapeptides and seven nonapeptides, with the amino acid compositions of the main repeated motifs highly similar to those of HMW glutenin subunits (Fig. 4C). However, Glu-1St1 repetitive domain was devoid of GQQ tripeptide (Fig. 4C), which is generally found in typical x-type subunits (Shewry et al., 1995). It also lacked the nonapeptide GHCPTSPQQ that is usually present towards the C-terminus of the repetitive domain of typical y-type subunits (Fig. 4C). Finally, the repetitive domain of Glu-1St1 differed strongly from that of D-hordein in both domain organization and repeated motifs. It lacked the non-repetitive and cysteine-containing fragment and the second repetitive region observed in D-hordein (Fig. 4C).

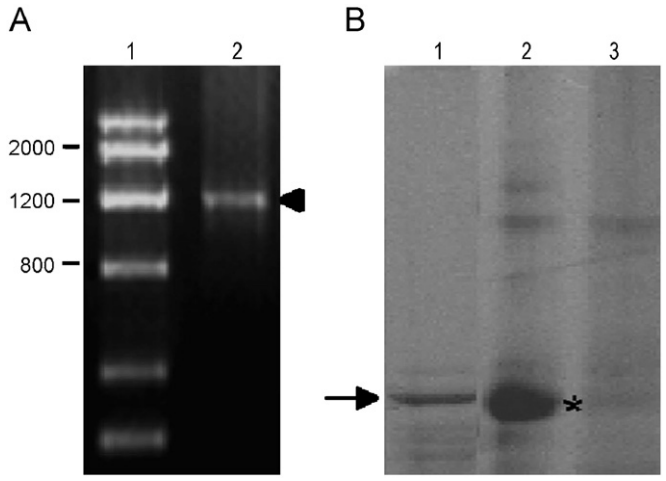

Fig. 2. Isolation of the complete genomic coding sequence of Glu-1St1 expressed in *Pd. stipifolia* accession PI325181. (A) A single and discrete PCR fragment (indicated by arrowhead in lane 2) was amplified from the genomic DNA sample of PI325181 using the primers (P1, P2) specific for the coding region of HMW glutenin subunit genes. The DNA markers (bp) were contained in lane 1. (B) Bacterial expression of the cloned Glu-1St1 coding sequence. The expression of Glu-1St1 mature protein (marked by asterisk) was detected in the IPTG induced bacterial culture (lane 2). This protein species was not observed in the control bacterial culture, which was not induced by IPTG (lane 3). The Glu-1St1 protein overexpressed in the bacterial cells showed an electrophoretic mobility identical to that of the native Glu-1St1 subunit extracted from the seed of *Pd. stipifolia* accession PI325181 (lane 1, indicated by arrow).

3.4. Phylogenetic relationships among Glu-1St1 and its orthologous subunits

The phylogenetic relationships of Glu-1St1 with its orthologous subunits from wheat and other *Triticeae* species were investigated by amino acid sequence comparisons and construction of phylogenetic trees. In the absence of a suitable outgroup to the topology tree, the two paralogous groups in the tree are used as outgroups for one another (Iwabe et al., 1989).

On the basis of the amino acid sequences of N- and C-terminal domains, two topology trees were constructed, each of which was divided into two halves consisting of Glu-1-1 and Glu-1-2 subunits, respectively (Fig. 5A and B). The topology tree constructed based on the N-terminal domain amino acid sequences showed that Glu-1St1 (Fig. 5A, arrowed) clustered with y-type HMW glutenin subunits, whereas that constructed using the C-terminal domain amino acid sequences revealed that Glu-1St1 (Fig. 5B, arrowed) formed a clade with x-type subunits. The D-hordein protein behaved as Glu-1St1 in the phylogenetic trees (Fig. 5A and B). The divergence times between *Glu-1St1* and its orthologous alleles from wheat and other *Triticeae* species were investigated. As shown in Table 2, the divergence times of *Glu-1St1* from its

MAKRLVLFAAVVLALVALTAA**EGGASGQLQCERELQESSLEVCRRVVDQQ**
LAGQLPWSTGLQMRCCQQLRDVSPECRPVAVSQVARQYEQQTAVPPKGGS
FYPGEATPPQQLQQRIFWGRSSQTVQGYYPSITSPQQGSYYPGQASPQQP
GQGQQQGQWQEPGQGQQPGQWKQPGQGQQGYYPTSPQQPGQGQQGYYPTS
PQQPGQGQQPGQWQRPGQGQQGYYPTSPQQPGQGQQPGQWQQPGQGQQGY
YPTSPQQPGQGQQPGQGQQPGQGQQGYYPTSPQQPGQGQQPGQWQQPGQG
QQPGQWQQPGQGQQGYYPTSSQQPGQGQQPGQWQQPGQGQQGYYPTSLQQ
PGQGQQSGQEQQGYG*SPYHVSAEQQVASLKVAKAQQLAAQLPAMCRLEGG*
DALSASQ (407 aa)

Fig. 3. The deduced amino acid sequence and primary structure of Glu-1St1. The amino acid sequence of Glu-1St1 contained 407 residues. The underlined sequence elements were confirmed by peptide motif fingerprinting of native Glu-1St1 subunit in MALDI-TOF-MS analysis. The N- and C-terminal domains are composed of the residues represented by bold and italic letters, respectively. The 21 residues in front of the N-terminal domain is the signal peptide. The central repetitive domain is located in between the N- and C-terminal domains.

Table 1
MALDI-TOF-MS analysis of peptide mass fingerprint of native Glu-1St1 subunit

Position	Missed cleavage	Calculated mass $(M+H)^+$	MALDI-MS measured $(M+H)^+$	Peptide sequence predicted using the bioinformatics program peptide mass (with trypsin digestion)[a]
1–12	0	1292.366	1291.694	EGGASGQLQCER
13–23	0	1350.486	1349.762	ELQESSLEVCR
25–43	0	2128.453	2127.223	VVDQQLAGQLPWSTGLQMR
50–65	0	1771.001	1770.015	DVSPECRPVAVSQVAR
66–76	0	1289.430	1287.629	QYEQQTAVPPK
77–94	0	1962.128	1963.020	GGSFYPGEATPPQQLQQR
77–99	1	2621.917	2622.366	GGSFYPGEATPPQQLQQRIFWGR
100–151	0	5680.999	5681.220	SSQTVQGYYPSITSPQQGSYYPG QASPQQPGQGQQQGQWQEPGQ GQQPGQWK
152–359[b]	0	22454.566	22457.21	QPGQGQQGYYPTSPQQPGQGQQ GYYPTSPQQPGQGQQPGQWQRP GQGQQGYYPTSPQQPGQGQQPGQWQQPG QGQQGYYPTSPQQPGQGQQPGQ GQQPGQGQQGYYPTSPQQPGQ GQQPGQWQQPGQGQQPG QWQQPGQGQQGYYPTSSQQ PGQGQQPGQWQQPG QGQQGYYPTSLQQPGQGQQS GQEQQGYGSPYHVSA EQQVASLK
360–375	1	1757.122	1756.180	VAKAQQLAAQLPAMCR
363–375	0	1458.736	1457.874	AQQLAAQLPAMCR
376–386	0	1048.094	1048.223	LEGGDALSASQ

[a]PeptideMass is available at http://www.expasy.ch/tools/peptide-mass.html.
[b]This peptide contains a predicted internal trypsin cleavage site (in between the underlined R and P residues). This site may not be cleaved efficiently by trypsin (Olsen et al., 2004).

orthologous alleles in common wheat (A, B, D), rye (R) and barley (H) were about 7.9 ± 1.1, 8.7 ± 1.1 and 9.5 ± 1.2 million years ago (MYA), respectively.

The unusual structural features displayed by Glu-1St1 also led us to investigate the 5′ flanking sequence of its coding region. By genomic PCR amplification, a 959 bp fragment upstream of the start codon of Glu-1St1 coding sequence was obtained. This region of *Glu-1St1* ORF contained an identical set of *cis*-elements found in the corresponding regions of previously characterized wheat HMW glutenin subunit genes (Shewry et al., 1999; Thomas and Flavell, 1990) (data not shown). Phylogenetic analysis showed that the 5′-flanking sequence of *Glu-1St1* clustered with those of x-type HMW glutenin subunit genes (data not shown).

4. Discussion

In the work described here, we investigated the HMW glutenin subunits in several *Pseudoroegneria* accessions, and analyzed the Glu-1St1 subunit from the *Pd. stipifolia* accession PI325181 in more detail. Based on the data depicted in Fig. 1, we conclude that HMW glutenin subunits are expressed in both diploid (*Pd. stipifolia*) and tetraploid (*Pd. tauri*, *Pd. strigosa*) *Pseudoroegneria* species. Importantly, the electrophoretic mobilities of the HMW glutenin subunits from the three *Pseudoroegneria* species examined in this work were all much faster than those of their orthologs from wheat. Further investigations are underway to examine if the above feature is common to the HMW glutenin subunits encoded by other *Pseudoroegneria* species. Another observation was that the numbers of HMW glutenin subunits expressed in the three *Pseudoroegneria* species appeared to differ widely, with the finding of only one subunit in the two accessions of the diploid species *Pd. stipifolia*. The expression of only one HMW glutenin subunit has previously been observed in *Triticum urartu* (Bai et al., 2004; D'Ovidio et al., 1996), which is diploid and the donor of the A genome of hexaploid wheat (Feldman, 2000). It has been shown that the *1Ay* gene is silenced in some *T. urartu* accessions, which leads to the expression of only x-type subunits in the seeds (Bai et al., 2004; D'Ovidio et al., 1996). Further studies are required to clarify if the expression of only one subunit in the diploid *Pd. stipifolia* accessions analyzed in this work is owing to the presence of a single HMW glutenin gene or the result of gene silencing.

To gain deeper insights into the HMW glutenin subunits of *Pseudoroegneria* species, we focused on analyzing the primary structure of Glu-1St1 expressed in the *Pd. stipifolia* accession PI325181. The genomic ORF of Glu-1St1 was cloned and found to be an authentic representation of the native *Glu-1St1* coding sequence by *in vitro* expression. The amino acid sequence deduced from the cloned ORF was judged to be identical to that of the native subunit by peptide motif fingerprinting using

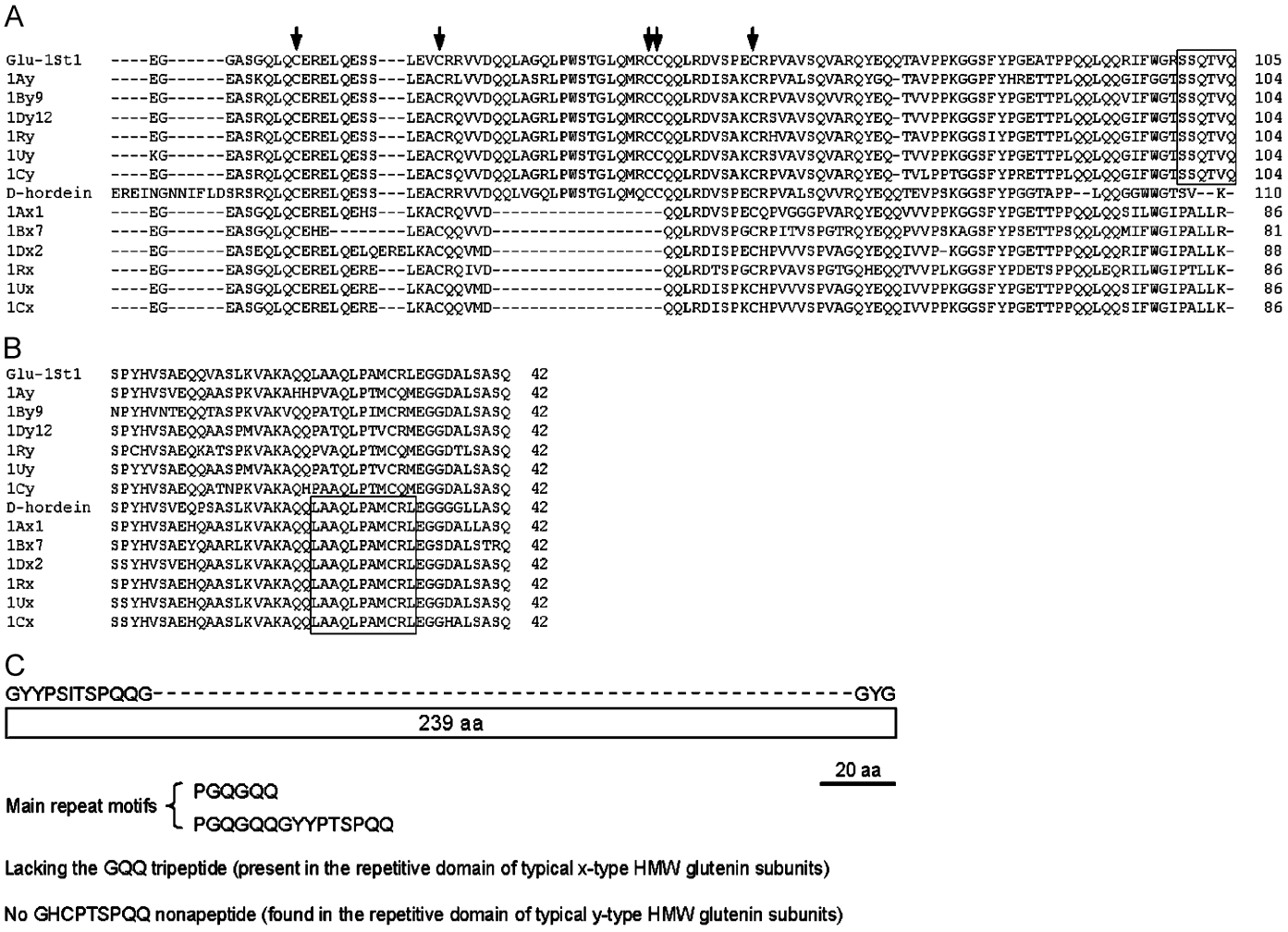

Fig. 4. Analysis of the structural features of Glu-1St1 domains. The analysis was carried out by amino acid sequence comparisons involving Glu-1St1, barley D-hordein, typical x-type HMW glutenin subunits (1Ax1, 1Bx7, 1Dx2, 1Rx, 1Ux, 1Cx) and typical y-type subunits (1Ay, 1By9, 1Dy12, 1Ry, 1Uy, 1Cy). (A) Glu-1St1 N-terminal domain exhibited higher similarity to those of y-type subunits and D-hordein based on the conservations in domain size, the number and positions of cysteine residues (marked by arrows) and the sequence element SSQTVQ (boxed region). (B) Glu-1St1 C-terminal domain displayed higher similarity to those of x-type subunits and D-hordein. The sequence element LAAQLPAMCRL (boxed region) was strictly conserved among the C-terminal domains of Glu-1St1, x-type subunits and D-hordein but not those of y-type subunits. (C) The size and organizations of the central repetitive domain of Glu-1St1. This domain started with the sequence element GYYPSITSPQQG and ended with the tripeptide GYG. The main repeat motifs in the domain were PGQGQQ and PGQGQQGYYPTSPQQ. It lacked the tripeptide GQQ (present in the repetitive domain of typical x-type subunits) and the sequence element GHCPTSPQQ (commonly found in the repetitive domain of typical y-type subunits).

MALDI-TOF-MS experiments. From amino acid sequence comparisons, we conclude that the domain organization in Glu-1St1 primary structure is identical to that of known HMW glutenin subunits. However, the length of Glu-1St1 repetitive domain was substantially shorter than that of typical HMW glutenin subunits, which explains the smaller size of Glu-1St1 and its much faster electrophoretic mobility in SDS-PAGE (Fig. 1). Our finding reinforces the suggestion that length variation of the repetitive domain may be a main factor causing size variations among different HMW glutenin subunit proteins (Shewry et al., 1995). From this point of view, it will be important to investigate if smaller repetitive domains may also be responsible for the reduced size of the HMW glutenin subunits in the *Pd. tauri* and *Pd. strigosa* accessions examined in this work in further experiments.

More detailed inspections of the amino acid sequences of Glu-1St1 domains revealed further unusual characteristics. Although the N-terminal domain of Glu-1St1 showed high similarity to that of typical y-type subunits, its C-terminal domain resembled that of typical x-type subunits. Its repetitive domain differed from typical x-type subunits by being devoid of the GQQ tripeptide and from typical y-type subunits by lacking the nonapeptide GHCPTSPQQ. Although the N- and C-terminal domains of Glu-1St1 shared conserved features with those of barley D-hordein, the repeat motifs and the organization of its repetitive domain were more similar to those of HMW glutenin subunits than to those of D-hordein. A closer relationship

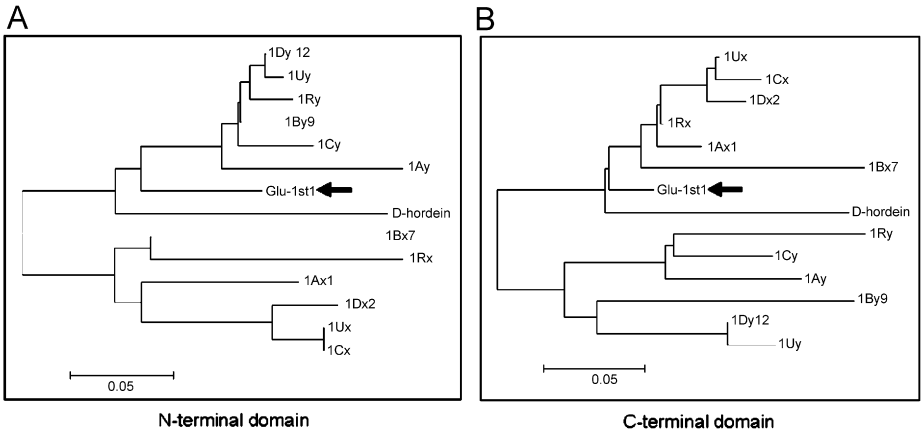

Fig. 5. Investigations of the phylogenetic relationships among Glu-1St1, its orthologous subunits from wheat (1Ax1, 1Bx7, 1Dx2, 1Ay, 1By9, 1Dy12), *Aegilops umbellulata* (1Ux, 1Uy), *Aegilops caudata* (1Cx, 1Cy) and rye (1Rx, 1Ry), and D-hordein from barley. The investigations were facilitated by the construction phylogenetic trees using the tree building methods installed in the MEGA website (http://www.megasoftware.net). The trees displayed in (A) and (B) were derived using the neighbor joining method, but identical results were obtained using the minimum evolution program. The distances among the compared domains were calculated using the pairwise deletion option and the amino acid substitution model *p*-distance. (A) A phylogenetic tree constructed using the N-terminal domain amino acid sequences. Glu-1St1 (arrowed) clustered with typical y-type subunits. (B) A phylogenetic tree constructed using the C-terminal domain amino acid sequences. Glu-1St1 (arrowed) aggregated with typical x-type subunits.

Table 2
Divergence times of *Glu-1St1* from orthologous HMW glutenin subunits genes in common wheat and rye and the *D-hordein* gene from barley

		Glu-1 allele	*Glu-R1* allele	*D-hordein*
Glu-1St1	Distance[a]	0.103±0.014	0.113±0.014	0.123±0.016
	MYA	7.9±1.1	8.7±1.1	9.5±1.2

[a]For calculating the divergence time, an average nucleotide substitution rate of 6.5×10^{-9} per site per year was used.

between Glu-1St1 and HMW glutenin subunits was also evident based on the higher similarity of *Glu-1St1* 5′-flnaking sequence to the corresponding regions of typical HMW glutenin subunit genes. Collectively, our analyses suggest that Glu-1St1 is a novel variant of HMW glutenin subunits. Considering that the divergence time between Glu-1St1 and its wheat orthologs was significantly shorter than that between Glu-1St1 and D-hordein, we speculate that Glu-1St1 may represent an evolutionary intermediate between D-hordein and typical HMW glutenin subunits. In this context, it becomes very important to investigate how widely the Glu-1St1-type subunits may distribute in *Triticeae* species, and whether the genes encoding Glu-1St1-type subunits have also been involved in gene duplication event (as what has been found for the genes encoding typical HMW glutenin subunits). These investigations should contribute to the elucidation of the mechanisms underlying the evolution of D-hordein and HMW glutenin subunits in *Triticeae* species.

Studies in bread wheat have indicated that the N- and C-terminals and the central repetitive domains all contribute to the function of HMW glutenin subunits in controlling dough quality (Field et al., 1987; Tatham et al., 1985, 1990). Because of the novelties exhibited by Glu-1St1 and its domains, it will be interesting to express this subunit in wheat endospermic tissues and to examine its potential effect on wheat dough properties. To this end, we are now conducting transgenic experiments to express the Glu-1St1 coding sequence in wheat endospermic cells.

Acknowledgements

This project is supported by grants from the Ministry of Science and Technology of China (2002CB111300 and 2006BAD13B02), the National Natural Science Foundation of China (30521001), and Chinese Academy of Sciences (CXLY-2002-6, KSCX2-SW-304). The authors are grateful to the constructive comments made by two anonymous reviewers.

References

Bai, J.-R., Jia, X., Liu, K.-F., Wang, D.-W., 2004. Cloning and characterization of the coding Sequences of the *1Ay* high molecular weight glutenin subunit genes from *Triticum urartu*. Acta Botanica Sinica 46, 463–471.

Cunsolo, V., Foti, S., Saletti, R., Gilbert, S., Tatham, A.S., Shewry, P.R., 2002. Investigation and correction of the gene-derived sequence of glutenin subunit 1Dx2 by matrix-assisted laser desorption/ionisation mass spectrometry. Rapid Communications Mass Spectrometry 16, 1911–1918.

Cunsolo, V., Foti, S., Saletti, R., Gilbert, S., Tatham, A.S., Shewry, P.R., 2003. Structural studies of glutenin subunits 1Dy10 and 1Dy12 by matrix-assisted laser desorption/ionisation mass spectrometry and high-performance liquid chromatography/electrospray ionisation mass spectrometry. Rapid Communications Mass Spectrometry 17, 442–454.

De Bustos, A., Rubio, P., Jouve, N., 2001. Characterization of two gene subunits on the 1R chromosome of rye as orthologs of each of the *Glu-1* genes of hexaploid wheat. Theoretical and Applied Genetics 103, 733–742.

De Bustos, A., Jouve, N., 2003. Characterization and analysis of new HMW-glutenin alleles encoded by the Glu-R1 locus of *Secale cereale*. Theoretical and Applied Genetics 107, 74–83.

Deng, Z., Zhang, X., 2004. Evolution of partial promoter region of HMW glutenin genes from E and E1 genome of *Agropyron elongatum*. Hereditas (Beijing) 26, 325–329.

D'Ovidio, R., Masci, S., Porceddu, E., 1996. Sequence analysis of the 5′ non-coding regions of active and inactive *1Ay* HMW glutenin genes from wild and cultivated wheats. Plant Science 114, 61–69.

Feldman, M., 2000. Origin of cultivated wheat. In: Bonjean, A.P., Angus, W.J. (Eds.), The World Wheat Book: A History of Wheat Breeding. Intercept Ltd, New York, pp. 3–56.

Field, J.M., Tatham, A.S., Shewry, P.R., 1987. The structure of a high Mr subunits of durum wheat (*Triticum durum*) glutenin. Journal of Biochemistry 247, 215–221.

Gaut, B.S., Morton, B.R., Mccaig, B.C., Clegg, M.T., 1996. Substitution rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene *Adh* parallel rate differences at the plastid gene *rbcL*. Proceedings of the National Academy of Sciences USA 93, 10274–10279.

Gu, Y.Q., Anderson, O.D., Londeore, C.F., Kong, X.Y., Chibbar, R.N., Lazo, G.R., 2003. The structural organization of the barley D-hordein locus in comparison with its orthologous regions of wheat genomes. Genome 46, 1084–1097.

Halford, N.G., Tatham, A.S., Sui, E., Daroda, L., Dreyer, T., Shewry, P.R., 1992. Identification of a novel beta-turn-rich repeat motif in the D hordeins of barley. Biochemistry Biophysica Acta 1122, 118–122.

Iwabe, N., Kuma, K.J., Hasegawa, M., Osawa, S., Miyata, T., 1989. Evolutionary relationships of archaebacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. Proceedings of the National Academy of Sciences USA 86, 9355–9359.

Jensen, K.B., 1990. Cytology and taxonomy of *E. grandiglumis*, *E. alatavicus*, and *E. batalinii* (Poaceae: Triticeae). Genome 33, 668–673.

Kumarathasan, P., Mohottalage, S., Goegan, P., Vincent, R., 2005. An optimized protein in-gel digest method for reliable proteome characterization by MALDI-TOF-MS analysis. Analytical Biochemistry 346, 85–89.

Lawrence, G.J., Shepherd, K.W., 1981. Chromosomal location of genes controlling seed protein in species related to wheat. Theoretical and Applied Genetics 59, 25–31.

Liu, Z., Yan, Z., Wan, Y., Liu, K., Zheng, Y., Wang, D., 2003. Analysis of HMW glutenin subunits and their coding sequences in two diploid *Aegilops* species. Theoretical and Applied Genetics 106, 1368–1378.

Li, W., Wan, Y., Liu, Z., Liu, K., Liu, X., Li, B., Li, Z., Zhang, X., Dong, Y., Wang, D., 2004. Molecular characterization of HMW glutenin subunit allele *1Bx14*: further insights into the evolution of *Glu-B1-1* alleles in wheat and related species. Theoretical and Applied Genetics 109, 1093–1104.

Lu, B.R., 1994. The genus *Elymus* L. in Asia. Taxonomy and biosystematics with special reference to genomic relationships. In: Wang, R.R.C., Jensen, K.B., Jaussi, C. (Eds.), Proceedings of the Second International *Triticeae* Symposium. Logan, UT, pp. 219–233.

McMillan, E., Sun, G., 2004. Genetic relationships of tetraploid *Elymus* species and their genomic donor species inferred from polymerase chain reaction-restriction length polymorphism analysis of chloroplast regions. Theoretical and Applied Genetics 108, 535–542.

Nei, N., Kumar, S., 2000. Molecular Evolution and Phylogenetics. Oxford University Press, Oxford.

Olsen, J.V., Ong, S.-E., Mann, M., 2004. Trypsin cleaves exclusively C-terminal to arginine and lysine residues. Molecular and Cellular Proteomics 3.6, 608–614.

Payne, P.I., 1987. Genetics of wheat storage protein and the effect of allelic variation on bread making quality. Annals Review of Plant Physiology 38, 141–153.

Payne, P.I., Lawrence, G.J., 1983. Detection of a novel high-molecular-weight subunit of glutenin in some Japanese hexaploid wheat. Journal of Cereal Science 1, 3–8.

Sanderson, M.J., 1998. Estimating rate and time in molecular phylogenies: beyond the molecular clock? In: Soltis, D.E., Soltis, P.S., Doyle, J.J. (Eds.), Molecular Systematics of Plants II: DNA Sequencing. Kluwer, Boston/Dordrecht/London, pp. 242–264.

Shewry, P.R., Halford, N.G., 2002. Cereal seed storage proteins: structures, properties and role in grain utilization. Journal of Experimental Botany 53, 947–958.

Shewry, P.R., Tatham, A.S., Barro, P., Lazzeri, P., 1995. Biotechnology of breadmaking: unraveling and manipulating the multi-protein gluten complex. Nature Biotechnology 13, 1185–1190.

Shewry, P.R., Tatham, A.S., Halford, N.G., 1999. The prolamins of the *Triticeae*. In: Shewry, P.R., Casey, R. (Eds.), Seed Proteins. Kluwer Academic Publishers, Dordrecht, pp. 35–78.

Sun, X., Hu, S., Liu, X., Qian, W., Hao, S., Zhang, A., Wang, D., 2006. Characterization of the HMW glutenin subunits from *Aegilops searsii* and identification of a novel variant HMW glutenin subunit. Theoretical and Applied Genetics 113, 631–641.

Tatham, A.S., Miflin, B.J., Shewry, P.R., 1985. The beta-turn conformation in wheat gluten proteins: relationship to gluten elasticity. Cereal Chemistry 62, 405–412.

Tatham, A.S., Drake, A.F., Shewry, P.R., 1990. Conformational studies of synthetic peptides corresponding to the repetitive regions of the HMW-GS of wheat. Journal of Cereal Science 11, 189–200.

Thomas, M.S., Flavell, R.B., 1990. Identification of an enhancer element for the endosperm-specific expression of high molecular weight glutenin. Plant Cell 2, 1171–1180.

Thompson, R.D., Bartrls, D., Harberd, N.P., Flavell, R.B., 1983. Characterization of the multigene family coding for HMW glutenin subunits in wheat using cDNA clones. Theoretical and Applied Genetics 67, 87–96.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research 22, 4673–4680.

Wan, Y., Wang, D., Shewry, P.R., Halford, N.G., 2002. Isolation and characterization of five novel high molecular weight subunit genes from *Triticum timopheevii* and *Aegilops cylindrica*. Theoretical and Applied Genetics 104, 828–839.

William, M.D., Peña, R.J., Mujeeb-Kazi, A., 1993. Seed protein and isozyme variations in *Triticum tauschii* (*Aegilops squarrosa*). Theoretical and Applied Genetics 87, 257–263.

Xu, D.H., Ban, T., 2004. Phylogenetic and evolutionary relationships between *Elymus humidus* and other *Elymus* species based on sequencing of non-coding regions of cpDNA and AFLP of nuclear DNA. Theoretical and Applied Genetics 108, 1443–1448.